

Uniform Resource Characteristics/Citations (URCs)

Environment of use

Documentation

A number of proposals and counter-proposals for URC formats have been made, usually by posting them as IETF Internet Drafts e.g.

- An SGML-based URC service, by Ron Daniel and Terry Allen
- Trivial URC syntax: urc0, by Paul Hoffman and Ron Daniel

In addition, various other formats have been mooted at one time or another as potential candidates for URCs. Since there have been a number of proposals, and no one clear favourite, this document will consider general aspects of the URC work and proposals, rather than concentrate on one particular URC proposal.

Documentation on the Uniform Resource Identifiers work can be found on the World-Wide Web at :

- <URL:<http://www.acl.lanl.gov/URI/>>
- <URL:<http://www.gatech.edu/iir/iir.html>>

Constituency of use

It is important to note that there is (currently) no URC *per se*. The term URC has generally been used to identify:

- long term cataloguing information pertaining primarily to on-line resources
- a standardised means of associating so-called *metadata*, or describing information, with objects - not necessarily for cataloguing purposes
- information used as part of the process of resolving a Uniform Resource Name (URN) to a URL or URLs
- information used by applications when selecting a particular instance of a resource from a number of possibilities, not necessarily as part of a URN lookup.

URCs started off life as the responsibility of the Internet Engineering Task Force's Uniform Resource Identifiers working group, which was chartered to investigate both URCs and Uniform Resource Names (URNs) - persistent location independent naming. In an unusual step for the IETF, the URI group was disbanded due to what was felt to be a lack of progress.

At the time of writing, an effort was under way to form a new IETF working group specifically addressing URC issues, and with a more focussed remit than the old URI group. Specifically: the new group would focus on developing a common carrier architecture which could be used to package various resource description formats, rather than attempting to standardise upon one particular preferred format.

Ease of creation

Proposals have concentrated on formats which are readily created and understood by both humans and computer programs - typically encoded as plain text. It has been assumed that specialist training would not be required for human beings, with the URC format typically being no more complex than an HTML document or the headers of an email message.

Progress towards international standardisation

Arguably, none. Some experimental implementations have been developed, but none has been widely deployed. This is not a pre-requisite for Internet protocol standardisation, but it is rare for a protocol to be standardised before it has been widely deployed.

Other comments

Despite the interest in long term cataloguing type information, most of the URC proposals which have emerged over the years have not addressed this - choosing instead to deal with simple technically oriented information such as the object's Internet Media type. A notable exception to this trend is the URC proposal, which attempts to address many of these considerations using an SGML DTD drawn from the Dublin Core work.

Format issues

Content

Basic descriptive elements

Typically a small number of attributes designed to contain information intended for automatic processing, e.g. selection between multiple replicas of a resource, or indexing by a Web Crawler type application. Some basic bibliographic details may be present typically in a simplistic form e.g. it may be possible to indicate an object's *author*, but not whether this is an institutional/corporate author, or an individual.

Subject description

This has not received much consideration, except within the SGML URC proposal.

URIs

All of the proposals deal with URIs explicitly, though in some circumstances it may be acceptable to have a URC which does not contain any URIs - e.g. when the resource is not available on-line.

Resource format and technical characteristics

Information about the resource format is typically provided using an Internet Media type. Some proposals also include other technical information such as size in bytes and transfer encoding.

Host administrative details

Not a major concern.

Administrative metadata

This is typically not present, though it may be possible to deduce by other means - e.g. HTTP headers.

Provenance/source

Not a major concern.

Terms of availability

Not a major concern.

Rules for the construction of these elements

Not a major concern.

Designation

Typically this takes the form of either attribute-value pairs, in the style of mail/news headers or whois++/IAFA templates, or SGML Document Type Definitions.

For example, in the *trivial URC scenario* referred to above, a URC for the popular Z Shell package could be written as:

```
=====
ftp://ftp.math.gatech.edu/pub/zsh
The Z-shell, a command interpreter
for many UNIX systems
which is freely available to
anyone with FTP access. Zsh is more
powerful than every other common
shell (sh, ksh, csh, tcsh and
bash) put together. The maintainer
is Richard Coleman,
zsh@math.gatech.edu
=====
ftp://ftp.sterling.com/zsh
A mirror site in the US
=====
ftp://ftp.cenatls.cena.dgac.fr/pub/shells/zsh
A mirror site in France
=====
ftp://mrrl.lut.ac.uk/zsh
A mirror site in the UK
```

Note the use of equals signs "=" as delimiters between instance information, and that the only information provided, aside from the URL, for each instance is a textual description - and even this is optional. In the trivial URC proposal, the ===== delimiters could be augmented with an Internet Media Type (MIME type) to indicate when an object was available in multiple formats. By contrast, the SGML URC proposal referred to above provides mechanisms for specifying additional semantics in the URC:

<urc>

<urn>urn:x-dns-2:shells.unix.computing.subjects.int:zsh</urn>

<author>Coleman, Richard</author>

<author type="email">zsh@math.gatech.edu</author>

<title>The Z-shell</title>

```
<subject scheme="abstract">
A command interpreter for many UNIX systems
which is freely available to anyone with FTP access. Zsh is more
powerful than every other common shell (sh, ksh, csh, tcsh and
bash) put together.
</subject>
```

```
<instance>
<coverage>Canonical distribution site</coverage>
<url>ftp://ftp.math.gatech.edu/pub/zsh</url>
</instance>
```

```
<instance>
<coverage>A mirror site in the US</coverage>
<url>ftp://ftp.sterling.com/zsh</url>
</instance>
```

```
<instance>
<coverage>A mirror site in France</coverage>
<url>ftp://ftp.cenatls.cena.dgac.fr/pub/shells/zsh</url>
</instance>
```

```
<instance>
<coverage>A mirror site in the UK</coverage>
<url>ftp://mrrl.lut.ac.uk/zsh</url>
</instance>
```

```
</urc>
```

In this case, parsing the URC is much more difficult, but there is the reward of being able to express complex relationships between objects within the URC framework.

Encoding

Human readable plain text encodings have been the norm for URC proposals. It should also be noted that most proposals have not made a distinction between the information being represented and its encoding, and have made no provision for multiple encodings of the same information.

Multi-lingual issues

Language and character set variants of an object have been considered in some of the URC proposals. Only the whois++ based scenarios appear to go any way towards addressing these issues when they arise within the URC itself e.g. when the abstract associated with a document-like object is available in multiple language or character set variants.

Ability to represent relationships between objects

Most URC proposals have effectively codified a small number of well known relationships, e.g. between URN and URL(s), between an object and its creator, and so on.

Fullness

Variable from minimal to rich, depending on the proposal selected. Most proposals err on the side of caution and use a minimal set of attributes.

Protocol issues

Some URC scenarios have been allied to particular protocols, e.g. whois++ and HTTP. HTTP seems to be of primary interest as a means of transporting URCs, which is understandable given the popularity it currently enjoys. Some protocols would not be particularly suited to shifting URCs around - for example, SGML URCs would need to be specially *packaged* for transport over whois++, since the protocol is optimised for attribute-value pairs.

The most likely scenario for the proposed IETF URC group would seem to be to register a top level Internet Media type for URCs (and/or metadata formats in general), under which various metadata formats could be registered. This would provide the necessary convention within the MIME framework for metadata formats to be transported in not just the World-Wide Web (via HTTP), but also in MIME enabled mail and news software. A sample application of this approach would be to provide machine readable announcements of new software packages, Web sites, and so on. It would also neatly sidestep the arguments over preferred metadata formats which have prevented any real progress from being made on URCs in the past. It should be noted that although URC development has not been particularly rapid, the drive to introduce parental control on the material available via the Internet has led to the formation of a number of URC style efforts, typically using metadata embedded within HTML documents or the HTTP protocol. Perhaps the most notable example of this approach is the Platform Independent Content Selection (PICS) work sponsored by the World-Wide Web Consortium. Whilst PICS is oriented towards censorship, the format used is not limited to this application.

It has been suggested from time to time that URC implementations should be capable of supporting searching, e.g. so that the URC associated with a particular URL can be determined. whois++ would appear to be the most popular candidate for this search capability, though other protocols including Z39.50 and X.500 have been suggested. A cut down version of X.500, known as the Lightweight Directory Access Protocol (LDAP - see RFC 1777) has recently been adopted by Netscape Communications Corporation, for use in their *Directory Server* product. Whilst this appears to be primarily aimed at White Pages type applications, such as discovering email addresses, their stated aim is to incorporate support for LDAP into the Netscape Navigator World-Wide Web browser. Such browser support, if handled carefully, would effectively make LDAP the protocol of choice for the search and retrieval of URC type information. However, it remains to be seen whether LDAP will be supported in the sort of open ended way which is needed for these applications.

Implementations

A number of experimental implementations of the various URC schemes have been developed - the WWW pages referred to at the start of this section contain pointers to them.

[Next](#)

[Table of Contents](#)

Page maintained by: [UKOLN Metadata Group](#)

Last updated: 10-Jun-1998